DOCUMENT RESUME

ED 419 005 TM 028 302

AUTHOR	van der Linden, Wim J.
TITLE	Multidimensional Adaptive Testing with a Minimum Error-Variance Criterion.
INSTITUTION	Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
REPORT NO	RR-97-03
PUB DATE	1997-00-00
NOTE	20p.
AVAILABLE FROM	Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Ability; *Adaptive Testing; Algorithms; *Computer Assisted Testing; Criteria; Foreign Countries; *Maximum Likelihood Statistics; Selection
IDENTIFIERS	*Multidimensionality (Tests)

ABSTRACT

The case of adaptive testing under a multidimensional logistic response model is addressed. An adaptive algorithm is proposed that minimizes the (asymptotic) variance of the maximum-likelihood (ML) estimator of a linear combination of abilities of interest. The item selection criterion is a simple expression in closed form. In addition, it is shown how the algorithm can be adapted if the interest is in a test with a "simple information structure." The statistical properties of the adaptive ML estimator are demonstrated for a two-dimensional item pool with several linear combinations of the two abilities. (Contains 1 figure and 15 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 419 005

TM

Multidimensional Adaptive Testing with a Minimum Error-Variance Criterion

Research Report 97-03

Wim J. van der Linden

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM028302

BEST COPY AVAILABLE

Multidimensional Adaptive Testing with a Minimum Error-Variance Criterion

Wim J. van der Linden

Abstract

The case of adaptive testing under a multidimensional logistic response model is addressed. An adaptive algorithm is proposed that minimizes the (asymptotic) variance of the maximum-likelihood estimator of a linear combination of abilities of interest. The item selection criterion is a simple expression in closed form. In addition, it is shown how the algorithm can be adapted if the interest is in a test with a "simple information structure". The statistical properties of the adaptive ML estimator are demonstrated for a two-dimensional item pool with several linear combinations of the two abilities.

Multidimensional Adaptive Testing
with a Minimum Error Variance Criterion

Adaptive testing algorithms for item pools calibrated under a unidimensional item response theory (IRT) model have been well investigated (e.g., Lord, 1980; Wainer, 1990), and several large-scale testing programs are in the process of introducing adaptive testing as an alternative to traditional paper-and-pencil tests. Since these programs need large item pools to guarantee measurement precision, in particular if measures to balance test content and control item exposure are implemented, violations of the assumption of unidimensionality of the item pool can be expected. Study of algorithms for adaptive testing under a multidimensional model seems therefore a timely matter.

The present paper is a sequel to van der Linden (1996) in which the problem of optimal assembly of a fixed test form from an item pool measuring multiple abilities is addressed. The emphasis in the work underlying this earlier paper was on an algorithm for assembling the fixed form to match optimally a set of targets for the (aymptotic) error variance functions for the abilities subject to a large variety of constraints on the composition of the test. It is the purpose of the present paper to study the use of the error variance as an item-selection criterion in tests with an adaptive format. Independent results on multidimensional adaptive testing are presented in Fan and Hsu (April, 1996) and Segall (1996). The interest in the former is in investigating the differences between item selection criteria based on various types of multivariate information measures rather than the error variance of the estimator(s). Also, these measures are evaluated over random sampling of correlated abilities. In the latter, the volumes of the confidence ellipsoid and the posterior credibility ellipsoid are proposed as multivariate item selection criteria. The posterior credibility ellipsoid is an attractive criterion because it allows for the possibility to build prior knowledge on dependencies between the ability variables into the item selection procedure. We will return to this point later in the paper.

The paper is organized as follows: The following section introduces the multidimensional IRT logistic model used in the presentation of the algorithm and motivates a linear combination of abilities as the parameter of interest in multidimensional adaptive testing. The subsequent section discusses the (asymptotic) variance of the estimator of a linear combination of ability parameters. Then it is proposed to minimize the variance of this estimator as a criterion for multidimensional adaptive testing, and an adaptive algorithm minimizing the variance is presented. The algorithm involves expressions of the item parameters which are easy to evaluate. The last section demonstrates the use of the algorithm for a two-dimensional item pool and investigates the statistical properties of the adaptive estimator for various linear combinations of the two abilities.

Multidimensional Model

Dichotomous response variables U_i are used to denote the responses of an examinee to item i=1,...,n. The variables take the value 1 if the response is correct and the value 0 if it is incorrect. The model is the following multivariate logistic response function:

$$p_i(\theta) \equiv \text{Prob}\{U_i = 1|\theta, a_i, d_i\} \equiv \frac{\exp(a_i'\theta - d_i)}{1+\exp(a_i'\theta - d_i)} \quad (1)$$

where $\theta=(\theta_1,...,\theta_j,...,\theta_m)$, with $-\infty < \theta_j < \infty$ for j=1,...,m, is a vector of m ability variables, $a_i=(a_{i1},...,a_{ij},...,a_{im})$, with $a_{ij}>0$ for j=1,...,m, is the vector of loadings of item i on these abilities (item discriminations), and $-\infty<d_i<\infty$ is a scalar representing a linear combination of the difficulties of the item along the ability dimensions. Detailed information about the model is given in McKinley and Reckase (1983), Reckase (1985, 1996), and Samejima (1974).

It is assumed that the item parameters a_i and d_i have already been estimated, and that the estimates are sufficiently accurate to consider them as the true parameter values. The parameters can be estimated using the Bayesian methods implemented in the program TESTFACT (Wilson, Wood, & Gibbons, 1984), or through McDonald's (1996) harmonic analysis applied to a normal approximation to the logistic function implemented in the program NOHARM (Fraser & McDonald, 1988).

Parameter of Interest

It is assumed that the parameter to be estimated by the adaptive testing procedure is a linear combination of the abilities, $\lambda'\theta$, where $\lambda=(\lambda_1,...,\lambda_m)$ is a vector of nonnegative weights. The choice of this parameter is motivated by the following practical cases:

1. The item pool is intentionally designed to measure more than one ability. However, the consumers of the test scores only want a single number to be reported. An obvious example is an item pool for a test to predict a future criterion of success in a selection problem, where the criterion is multifaceted. In this case, the weights λ_j are to reflect the relative importance of the individual abilities with respect to the criterion.
2. The item pool is designed to measure only one ability but the items are sensitive to some "nuisance abilities" as well. A well-know example is a test for mathematical ability depending on verbal abilities required to understand the items. This case can

BEST COPY AVAILABLE

be dealt with by setting the weights $\lambda_j=0$ for all nuisance abilities. As will become clear below, this measure does not neutralize the effect of the values of the nuisance parameters on the variance of the estimator of the intended ability but does allow for direct minimization of this variance.

3. Even though the item pool measures several abilities, different sections of the test may be required to be maximally informative with respect different abilities, for example, because identifiable subsections of the tests are to be used for diagnosing individual abilities. As will be shown below, an adaptive test with a "simple ability structure" over the ability space can be realized by choosing different values for the weights λ_j at different stages of the procedure. Note that this case is not equivalent to the one of choosing items from different unidimensional item pools; rather than rotating the selection of the items across unidimensional pools the weights λ_j are rotated across different preselected values while selecting items from a multidimensional pool.

For an extended description of the above cases of multidimensional testing using the format of a fixed-form test, see van der Linden (1996).

Ability Estimation

It is assumed that $\lambda'\theta$ is estimated by the method of MLE. As is well known, for MLE it holds that

$$\lambda'\theta = \lambda'\hat{\theta}. \tag{2}$$

For a response pattern $(u_1,...,u_n)$, the likelihood of θ is defined as

$$L(\theta; u_1,...,u_n,\mathbf{a}_1,...,\mathbf{a}_n,\mathbf{d}_1,...,\mathbf{d}_n) = \prod_{i=1}^{n} \text{Prob}\{U_i = u_i|\theta,\mathbf{a}_i,\mathbf{d}_i\}. \tag{3}$$

The joint MLE of θ_j, j=1,...,m, is the vector of values maximizing this likelihood. The likelihood equations are obtained setting the partial derivatives of the log of (3) equal to zero:

$$\frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^{n} u_i \ln p_i(\theta) + (1-u_i)\ln(1-p_i(\theta)) = 0, \quad j=1,...,m, \tag{4}$$

Using

$$\frac{\partial L}{\partial \theta_j} = a_{ij} P_i(\theta)[1 - P_i(\theta)], \tag{5}$$

the likelihood equations can be written as

$$\sum_{i=1}^{n} a_{ij}[u_i - P_i(\theta)] = 0, \quad j=1,....,m,. \tag{6}$$

which is the common form known to exist for a model belonging to the exponential family (Andersen, 1980, sect. 3.2). The system can be solved using Newton's method

$$\theta^{(t)} = \theta^{(t-1)} - [H(\theta^{(t-1)})]^{-1} \gamma(\theta^{(t-1)}), \tag{7}$$

where $H(\theta^{(t-1)})$ is the Hessian of the log-likelihood function with elements

$$\frac{\partial^2 \ln L}{\partial \theta_g \partial \theta_h} = -\sum_{i=1}^{n} a_{ig} a_{ih} P_i(\theta)[1 - P_i(\theta)], \qquad g,h = 1,....,m. \tag{8}$$

and $\gamma(\theta^{(t-1)})$ is the gradient of the log-likelihood function, i.e., the vector with the first derivatives set equal to zero in (6), both evaluated at step t-1. Substitution of the results from (7) into (2) gives the MLE of $\lambda'\theta$.

Variance of $\lambda'\theta$

The asymptotic covariance matrix of the MLE of θ is given by the inverse of Fisher's information matrix

$$I(\theta) = (-E\frac{\partial^2 \ln L(\theta; U_1,....,U_n, a_1,....,a_n, d_1,....,d_n)}{\partial \theta_g \partial \theta_h}), \tag{9}$$

with $L\theta; U_1,....,U_n, a_1,....,a_n, d_1,....,d_n)$ being the likelihood statistics associated with the random response vector and θ_p and θ_q any two components of θ. From (8), it follows that

$$-E\frac{\partial^2 \ln L(\theta; U_1,\ldots,U_n,a_1,\ldots,a_n,d_1,\ldots,d_n)}{\partial\theta_g\partial\theta_h} = \sum_{i=1}^{n} a_{ig}a_{ih}P_i(\theta)[1-P_i(\theta)]. \qquad (10)$$

Standard techniques for matrix inversion yield the (asymptotic) covariance matrix

$$\mathbf{V} \equiv \mathrm{Var}(\hat{\theta}|\theta) = \mathrm{I}(\theta)^{-1}, \qquad (11)$$

where the determinant of the information matrix, $|\mathrm{I}(\theta)|$, is assumed to be not vanishing. For the linear combination $\lambda'\hat{\theta}$

$$\mathrm{Var}(\lambda'\hat{\theta}|\theta) = \lambda'\mathrm{V}\lambda. \qquad (12)$$

For a model with two ability parameters, $\theta \equiv (\theta_1,\theta_2)$, and $(\lambda_1,\lambda_2) \equiv (\lambda,1-\lambda)$, the result in (12) boils down to

$$\begin{aligned}
&\mathrm{Var}(\lambda\hat{\theta}_1 + (1-\lambda)\hat{\theta}_2|\theta_1,\theta_2)\\
&= \lambda^2\mathrm{Var}(\hat{\theta}_1|\theta_1,\theta_2) + (1-\lambda)^2\mathrm{Var}(\hat{\theta}_2|\theta_1,\theta_2) + 2\lambda(1-\lambda)\mathrm{Cov}(\hat{\theta}_1,\hat{\theta}_2|\theta_1,\theta_2)\\
&= [\lambda^2\sum_{i=1}^{n} a_{i2}^2P_i(\theta_1,\theta_2)\{1-P_i(\theta_1,\theta_2)\} + (1-\lambda)^2\sum_{i=1}^{n} a_{i1}^2P_i(\theta_1,\theta_2)\{1-P_i(\theta_1,\theta_2)\}\\
&\qquad + 2\lambda(1-\lambda)\sum_{i=1}^{n} a_{i1}a_{i2}P_i(\theta_1,\theta_2)\{1-P_i(\theta_1,\theta_2)\}]/|\mathrm{I}(\theta_1,\theta_2)|, \qquad (13)
\end{aligned}$$

where

$$\begin{aligned}
|\mathrm{I}(\theta_1,\theta_2)| &= [\sum_{i=1}^{n} a_{i1}^2P_i(\theta_1,\theta_2)\{1-P_i(\theta_1,\theta_2)\}][\sum_{i=1}^{n} a_{i2}^2P_i(\theta_1,\theta_2)\{1-P_i(\theta_1,\theta_2)\}]\\
&\quad - [\sum_{i=1}^{n} a_{i1}a_{i2}P_i(\theta_1,\theta_2)\{1-P_i(\theta_1,\theta_2)\}]^2. \qquad (14)
\end{aligned}$$

Note that for n=2 the two items should not be parallel, i.e., it should not hold that

$$\begin{aligned}
a_{11}&=a_{21},\\
a_{12}&=a_{22},\\
d_1&=d_2,
\end{aligned} \qquad (15)$$

because then the determinant in (13) vanishes.

Adaptive Testing Algorithm

For notational convenience, the adaptive testing algorithm is also presented for the case of two ability parameters. The following definitions are needed: The items in the pool are indexed by i=1,...,I. The adaptive testing procedure is assumed to be stopped after n items have been administered. The order of the items in the test is indexed by k=1,...,n. Thus, i_k is the index of the item in the pool administered as the kth item in the test. Suppose k-1 items have been selected. Let $S_k=\{i_1,...,i_{k-1}\}$ denote this set of items. Then, $R_k=\{1,...,I\}\backslash S_k$ is the set of items rejected so far, and item i_k has to be selected from this set. Finally, let $\hat{\theta}_1^k$ and $\hat{\theta}_2^k$ be the estimators of θ_1 and θ_2 after k items have been administered.

The kth item is selected according to the following criterion:

$$\min_{R_k}\{\mathrm{Var}(\lambda\hat{\theta}_1^k + (1-\lambda)\hat{\theta}_2^k|\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\}, \tag{16}$$

that is, the item is selected to minimize the variance of $\lambda\hat{\theta}_1^k+(1-\lambda)\hat{\theta}_2^k$ evaluated at the current estimates, which is (13) for $(\theta_1,\theta_2)=(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})$.

To implement the criterion, define

$$u^i \equiv \sum_{g=1}^{k-1} a_{i_g1}^2 P_{i_g}(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\{1-P_{i_g}(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\} + a_{i1}^2 P_i(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\{1-P_i(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\}, \tag{17}$$

$$v^i \equiv \sum_{g=1}^{k-1} a_{i_g2}^2 P_{i_g}(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\{1-P_{i_g}(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\} + a_{i2}^2 P_i(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\{1-P_i(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\}, \tag{18}$$

$$w^i \equiv \sum_{g=1}^{k-1} a_{i_g1}a_{i_g2} P_{i_g}(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\{1-P_{i_g}(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\} + a_{i1}a_{i2} P_i(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\{1-P_i(\hat{\theta}_1^{k-1},\hat{\theta}_2^{k-1})\}. \tag{19}$$

The criterion can thus be expressed in closed form as

$$i_k = \min_j\{[\lambda^2 v^j+(1-\lambda)^2 u^j-2\lambda(1-\lambda)w^j]/[u^j v^j-(w^j)^2];\ j\in R_k\}\ . \qquad (20)$$

In order to select the next item, for each item in R_k one term is added to the sums in (17)-(19). These term involves both the parameters a_{j1}, a_{j2}, and d_j and the probability $P_j(\theta_1,\theta_2)$, where the last quantity is evaluated at the current estimates $\theta_1=\hat{\theta}_1^{k-1}$ and $\theta_2=\hat{\theta}_2^{k-1}$. The item minimizing the expression in (20) is selected.

Algorithm. The algorithm can be summarized as follows:

1. Choose a value for λ reflecting the validity of the test;
2. Select item i_1 and i_2 according to some external criterion;
3. Estimate θ_1 and θ_2 using the MLEs from (7);
4. Enter the values of the parameters of item i_1 and i_2 into (17)-(19);
5. Evaluate (17)-(19) for $i\in R_2$, and select the minimizer of (20);
6. Repeat Steps 3-5 for k=3,...,n.

In selecting the items in Step 2, the item parameters should be avoided to approximate the conditions in (15) because of instability in (20). As long as the examinee produces responses which are uniformly correct or incorrect, the MLEs of θ_j have to be bounded by well-chosen value.

Simple Structure. As already noted, different sections of the test may be required to have good measurement qualities with respect to different abilities ("simple ability structure"). For two abilities, let n_1 and n_2 be the required numbers of items in the two sections. The best method to obtain error variances of $\hat{\theta}_1$ and $\hat{\theta}_2$ proportional to the ratio of n_1 to n_2 seems to set λ equal to 1 n_1 times and to 0 n_2 times while alternating between the two sections from the beginning of the test seems. It should be noted, however, that with a multidimensional item pool the responses to each item contribute to the variance of $\hat{\theta}_1$ as well as $\hat{\theta}_2$, and therefore both variances must be calculated over all n_1+n_2 items in the test. Therefore, either variance may become more favorable than strictly required.

Numerical Examples

A pool of 500 items was simulated drawing random values for the parameters a_{j1} and a_{j2} from U(0.0,1.3) and for d_j from U(-1.3,1.3). The ranges of these distributions correspond roughly to the ranges of the parameter values in a two-dimensional ACT Assessment Program

BEST COPY AVAILABLE

Mathematics Item Pool used in van der Linden (1966) to study the performance of a linear programming model for assembling fixed-form tests. The adaptive algorithm was applied to simulated responses to a 50-item test of examinees with abilities on a two-dimensional grid defined by θ_1,θ_2 =-2.0, -1.8, ..., 2.0. Because the model in (1) permits MLE only when at least three responses are available, the full adaptive procedure was started only after responses to the first three items were simulated. The first two items were defined to have the same parameter values (a_{1i},a_{2i},d_i)=(1.2,0.1,0.0) and (0.1,1.2,0.0) for all examinees. The third item was selected from the pool applying (13) to simulated responses on the first two items. The log-likelihood function for the model in (1) is known to have an occasional unbounded maximum. In such cases, which happened predominantly for the combination of short test lengths and extreme values of (θ_1,θ_2), the ability estimates were truncated at ±2. For each combination of ability values, 100 replications were produced. The study was repeated for λ =0.250, 0.375, and 0.500 (larger values of λ were omitted because of symmetry).

Figure 1 shows the estimated bias and mean-squared error (MSE) of $\lambda\hat{\theta}_1^k+(1-\lambda)\hat{\theta}_2^k$ as a function of θ_1 and θ_2 for k=10, 30, and 50 items and the different values of λ. The dominant impression from these plots is that test length is a decisive factor but the choice of a value for weight λ hardly has any effect on the behavior of the estimator. At 10 items the estimator has a unfavorable MSE for all values of $\lambda\theta_1+(1-\lambda)\theta_2$. At the extremes, the large MSE is in part due to an intolerably large bias in the estimator. However, for 30 items the procedure seems to work reasonably well, and at 50 items both the MSE and the bias seem to be ignorable for all practical purposes. The fact that the results are robust with respect to λ implies that from a statistical point of view the size of λ is hardly important when setting up a program of multidimensional adaptive testing and that this factor can vary freely across applications without having any impact on the statistical properties of the ability estimator.

Conclusions

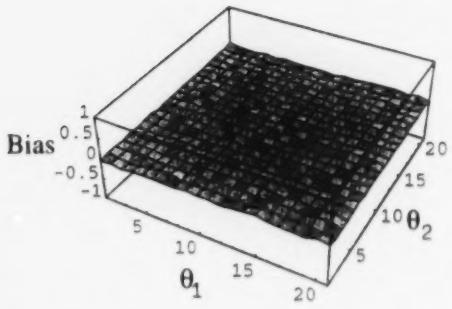
The main conclusion from this study is that use of the procedure in this paper seems practically feasible, provided the number of items in the test is set not too small. For short tests, the ML estimators of θ_1 and θ_2 are strongly biased and unstable, even when combined into a linear combination as in this paper. In this case, it seems better to resort to a Bayesian procedure as the one in Segall (1996). If empirical information about the correlation between the ability variables is available from external sources, a Bayesian procedure allows for the possibility of building this information into the (multivariate) prior distribution for the abilities. As a consequence, the adaptive estimator can be expected to stabilize quicker as a function of the test length.

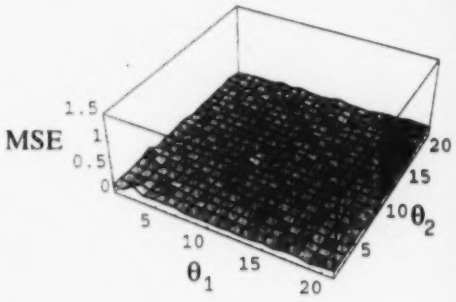
Figure 1.

Bias and MSE functions of $\lambda\theta_1^k + (1-\lambda)\theta_2^k$ for λ =0.250, 0.375, 0.500 (k=10, 30, 50)

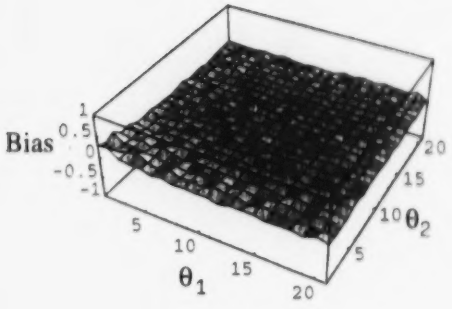
$\lambda = .250$

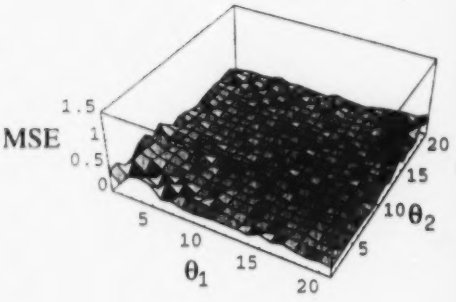
k = 50



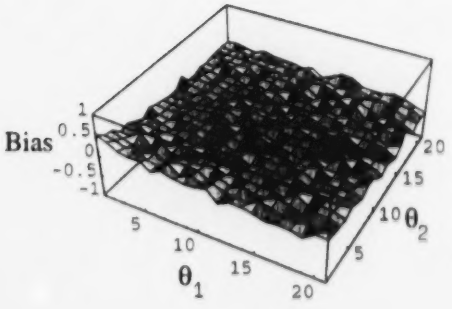


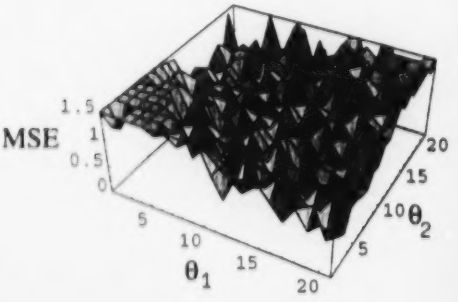
k = 30





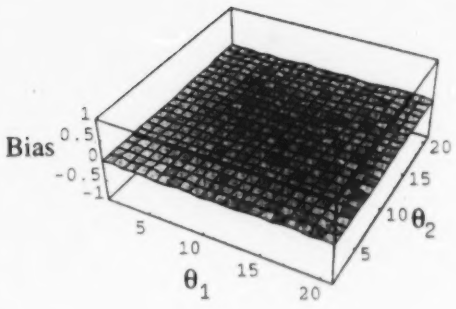
k = 10

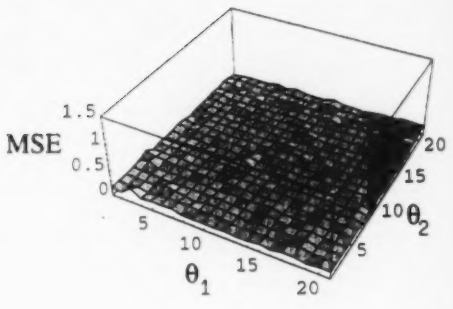




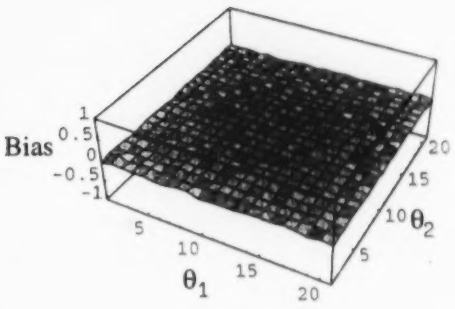
$\lambda = .375$

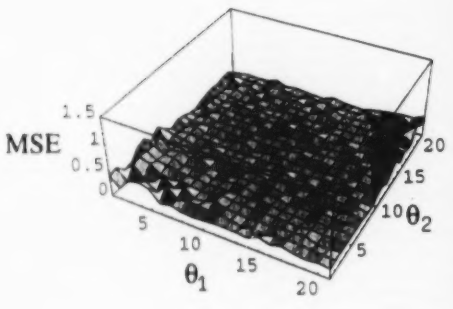
k = 50



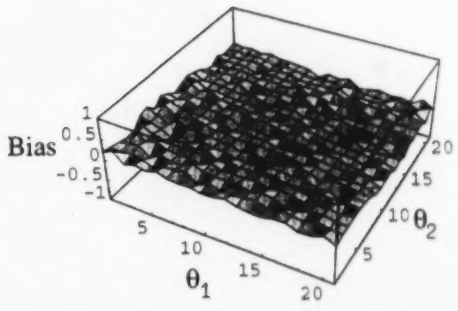


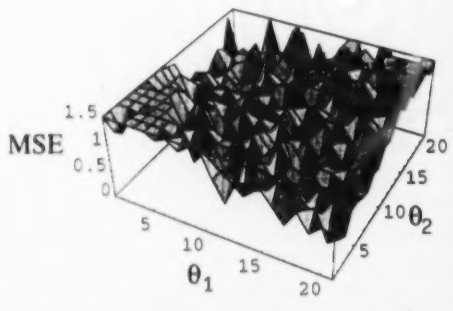
k = 30





k = 10

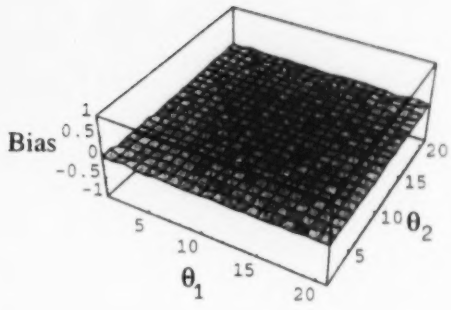


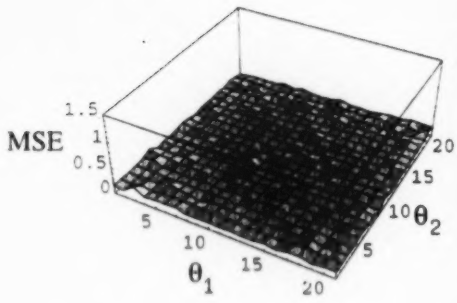


BEST COPY AVAILABLE

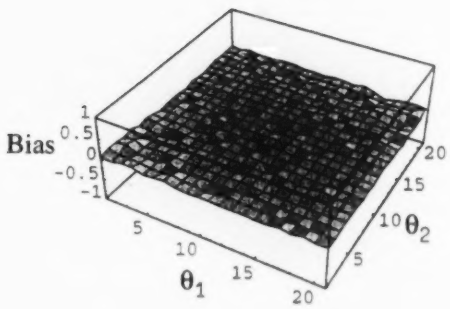
$\lambda = .500$

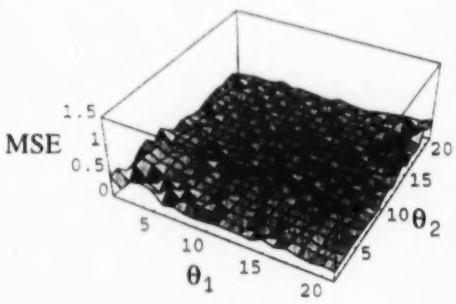
k = 50



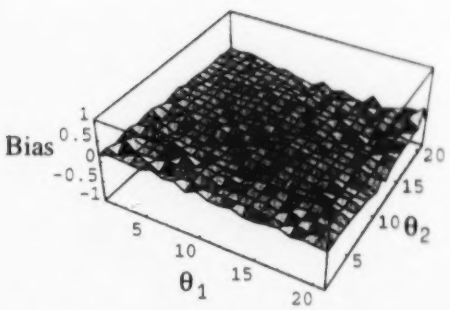


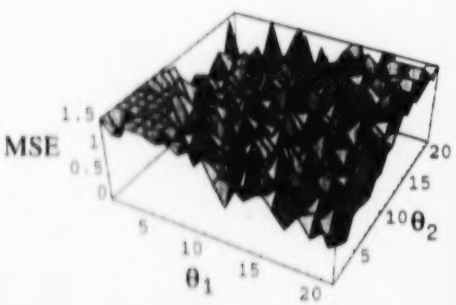
k = 30





k = 10





Multidimensional IRT has a tradition of using a multivariate information measure as a tool for analyzing a test searching the ability space for directions in which information is maximal. These directions then define the ability composite which the test is assumed to measure best (Ackerman, 1994; Fan, 1996; Reckase & McKinley, 1991). The orientation in the current paper has been different. The composite was defined to be the parameter of interest. Next, the adaptive algorithm was used to have uniform measurement accuracy across the ability space. As demonstrated by the flat MSE functions for k=30 and 50 in Figure 1, it is possible indeed to have the same favorable measurement precision for all ability points and not only for those points that "define the composite".

References

Ackerman, T.A. (1994). Creating a test information profile for a two-dimensional latent space. Applied Psychological Measurement, 18, 257-275.

Andersen, E.B. (1980). Discrete statistical models with social science applications. Amsterdam: North-Holland.

Fan, M., & Hsu, Y. (April, 1996). Multidimensional computer adaptive testing. Paper presented at the annual meeting of the American Educational Testing Association, New York City, NY.

Fraser, C. & McDonald, R.P (1988). NOHARM II: A FORTRAN program for fitting unidimensional and multidimensional normal ogive models of latent trait theory. Armindale, Australia: Centre for Behavioral Studies, University of New England.

Lord, F.M. (1980). Applications if item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

McDonald, R.P. (1966). Normal-ogive multidimensional model. In W.J. van der Linden & R.K. Hambleton (Eds.), Handbook of modern item response theory. New York City, NY: Springer-Verlag.

McKinley, R.L., & Reckase, M.N. (1983). An extension of the two-parameter logistic model to the multidimensional latent space (Research report ONR 83-2). Iowa City, IA: American College Testing.

Reckase, M.D. (1985). The difficulty of test items that measure more than one ability. Applied Psychological Measurement, 9, 401-412.

Reckase, M.D. (1966). A linear logistic multidimensional model for dichotomous item response data. In W.J. van der Linden & R.K. Hambleton (Eds.), Handbook of modern item response theory (pp. 271-286). New York City, NY: Springer-Verlag.

Reckase, M.D., & McKinley, R.L. (1991). The discriminating power of items that measure more than one dimension. Applied Psychological Measurement, 15, 361-374.

Samejima, F. (1974). Normal ogive model for the continuous response level in the multidimensional latent space. Psychometrika, 39, 111-121.

Segall, D.O. (1996). Multidimensional adaptive testing. Psychometrika, 61, 331-335.

van der Linden, W.J. (1996). Assembling tests for the measurement of multiple traits. Applied Psychological Measurement, 20, 373-388.

Wainer, H. (Ed.) (1990). Computerized adaptive testing: A primer. Hillsdale, NJ: Erlbaum.

Wilson, D. Wood, R, & Gibbons, R. (1984) TESTFACT: Test scoring, item statistics, and item factor analysis. Mooresville, IN: Scientific Software.

 BEST COPY AVAILABLE

Author Note

The author is most indebted to Wim M.M. Tielen for his computational support in preparing the numerical examples and to Cees A.W. Glas for his comments on an earlier version of this paper. Send all correspondence to: W.J. van der Linden, Department of Educational Measurement and Data Analyses, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. Email: vanderlinden@edte.utwente.nl.

Titles of Recent Research Reports from the Department of Educational Measurement and Data Analysis. University of Twente, Enschede, The Netherlands.

RR-97-03 W.J. van der Linden, *Multidimensional Adaptive Yesting with a Minimum Error-Variance Criterion*

RR-97-02 W.J. van der Linden, *A Procedure for Empirical Initialization of Adaptive Testing Algorithms*

RR-97-01 W.J. van der Linden & Lynda M. Reese, *A Model for Optimal Constrained Adaptive Testing*

RR-96-04 C.A.W. Glas & A.A. Béguin, *Appropriateness of IRT Observed Score Equating*

RR-96-03 C.A.W. Glas, *Testing the Generalized Partial Credit Model*

RR-96-02 C.A.W. Glas, *Detection of Differential Item Functioning using Lagrange Multiplier Tests*

RR-96-01 W.J. van der Linden, *Bayesian Item Selection Criteria for Adaptive Testing*

RR-95-03 W.J. van der Linden, *Assembling Tests for the Measurement of Multiple Abilities*

RR-95-02 W.J. van der Linden, *Stochastic Order in Dichotomous Item Response Models for Fixed Tests, Adaptive Tests, or Multiple Abilities*

RR-95-01 W.J. van der Linden, *Some decision theory for course placement*

RR-94-17 H.J. Vos, *A compensatory model for simultaneously setting cutting scores for selection-placement-mastery decisions*

RR-94-16 H.J. Vos, *Applications of Bayesian decision theory to intelligent tutoring systems*

RR-94-15 H.J. Vos, *An intelligent tutoring system for classifying students into Instructional treatments with mastery scores*

RR-94-13 W.J.J. Veerkamp & M.P.F. Berger, *A simple and fast item selection procedure for adaptive testing*

RR-94-12 R.R. Meijer, *Nonparametric and group-based person-fit statistics: A validity study and an empirical example*

RR-94-10 W.J. van der Linden & M.A. Zwarts, *Robustness of judgments in evaluation research*

RR-94-9 L.M.W. Akkermans, *Monte Carlo estimation of the conditional Rasch model*

RR-94-8 R.R. Meijer & K. Sijtsma, *Detection of aberrant item score patterns: A review of recent developments*

RR-94-7 W.J. van der Linden & R.M. Luecht, *An optimization model for test assembly to match observed-score distributions*

RR-94-6 W.J.J. Veerkamp & M.P.F. Berger, *Some new item selection criteria for adaptive testing*

RR-94-5 R.R. Meijer, K. Sijtsma & I.W. Molenaar, *Reliability estimation for single dichotomous items*

BEST COPY AVAILABLE

RR-94-4 M.P.F. Berger & W.J.J. Veerkamp, *A review of selection methods for optimal design*

RR-94-3 W.J. van der Linden, *A conceptual analysis of standard setting in large-scale assessments*

RR-94-2 W.J. van der Linden & H.J. Vos, *A compensatory approach to optimal selection with mastery scores*

RR-94-1 R.R. Meijer, *The influence of the presence of deviant item score patterns on the power of a person-fit statistic*

RR-93-1 P. Westers & H. Kelderman, *Generalizations of the Solution-Error Response-Error Model*

RR-91-1 H. Kelderman, *Computing Maximum Likelihood Estimates of Loglinear Models from Marginal Sums with Special Attention to Loglinear Item Response Theory*

RR-90-8 M.P.F. Berger & D.L. Knol, *On the Assessment of Dimensionality in Multidimensional Item Response Theory Models*

RR-90-7 E. Boekkooi-Timminga, *A Method for Designing IRT-based Item Banks*

RR-90-6 J.J. Adema, *The Construction of Weakly Parallel Tests by Mathematical Programming*

RR-90-5 J.J. Adema, *A Revised Simplex Method for Test Construction Problems*

RR-90-4 J.J. Adema, *Methods and Models for the Construction of Weakly Parallel Tests*

RR-90-2 H. Tobi, *Item Response Theory at subject- and group-level*

RR-90-1 P. Westers & H. Kelderman, *Differential item functioning in multiple choice items*

Research Reports can be obtained at costs, Faculty of Educational Science and Technology, University of Twente, Mr. J.M.J. Nelissen, P.O. Box 217, 7500 AE Enschede, The Netherlands.